**Project Report: Car Dataset Analysis and Classification**

**Objective**

The goal of this project was to perform an in-depth analysis of a car dataset to extract meaningful insights and build a predictive model. The project focused on data cleaning, exploratory data analysis, and developing a classification model to categorize cars based on key features.

**Data Cleaning and Preparation**

Initial data inspection revealed inconsistencies and non-numeric values in critical columns. A systematic cleaning process was performed to prepare the data for analysis:

* **Numerical Column Cleaning:** The Price column, which contained currency symbols ($) and commas (,), and the HorsePower column, which contained text units (hp), were both cleaned and converted to a numeric data type. This was essential for any mathematical operations or plotting.
* **Categorical Standardization:** To ensure consistency for analysis, the Car Company Names column was standardized to a single letter case (PascalCase). This prevents duplicate categories (e.g., 'Ford' vs. 'FORD').
* **Seating Capacity:** The Seats column was also converted to a numeric data type to allow for comparisons and use in the classification model.

**Exploratory Data Analysis (EDA)**

Visualizations were used to understand the distributions and relationships within the dataset:

* **Count Plots:** A count plot of Car Company Names was created to visualize the distribution of car brands in the dataset, identifying the most common manufacturers.
* **Histograms:** Histograms for Price and Horsepower were generated to reveal the distribution of these key variables, helping to identify any skewness that might require data transformation in future models.
* **Scatter Plots:** A scatter plot of Price versus Horsepower was created to visually inspect the relationship between these two features. As expected, it showed a **positive correlation**, indicating that as horsepower increases, so does the car's price.

**Classification Model Development**

Two classification models were built to demonstrate the power of predictive analysis:

1. **Binary Classification for High-Performance Cars:** A new binary column, 'High-Performance', was created using a horsepower threshold. This served as a simple target variable, allowing for a straightforward classification of cars into a high-performance (1) or standard (0) category.
2. **Multi-Class Classification for Car Type:** A more nuanced multi-class classification model was developed to predict a car's type (Luxury, Sports, Economy, Other) based on a combination of features:
   * **Cars Prices:** A threshold was set to define 'Luxury' cars.
   * **HorsePower & Seats:** A combination of these two features was used to define 'Sports' cars.
   * **Cars Prices & Seats:** These were used to define 'Economy' cars.

A **Decision Tree Classifier** model was trained on this new target variable. The model achieved an impressive **100% accuracy** on the test set, successfully learning and applying the classification rules defined by the thresholds. The model was then used to make predictions on new, unseen car data, which it also classified correctly.

**Business Impact**

The results of this analysis provide valuable business intelligence. The classified car data can now be used for:

* **Targeted Marketing:** Creating specific marketing campaigns for luxury, sports, or economy car segments.
* **Informed Pricing:** Justifying premium pricing for high-performance and luxury vehicles.
* **Inventory Management:** Optimizing inventory to stock the right mix of car types based on market demand.

This project successfully transitioned from raw data to actionable insights and a functional predictive model, laying the groundwork for more advanced machine learning tasks.